How to Make Best Use of the Intellectual Output of the Country? A Simple Approach to the Design of a Digital Library of Theses and Dissertations in Indian Universities

Dr. Gobinda G. Chowdhury
Department of Information Science
University of Strathclyde
Glasgow G1 1XH, UK
e-mail: gobinda@dis.strath.ac.uk

Introduction

Research and development activities in digital libraries are less than a decade old and virtually large-scale systematic research in digital libraries began with the Digital Libray Research Initiatives in the US in 1994 [1,2] and the Electronic Libraries programme (eLib) in UK in 1995 [3]. Nevertheless, Digital Library Initiative Phase 1 and Phase 2 [1,2] projects in the US and *eLib* phase 1, phase 2, and phase 3 projects [3] in UK have played a key role in digital library developments. At the moment there are many on-going digital library projects in different parts of the world and features of many of these have appeared in the literature (see for example, [4,5,6]). These digital library research projects introduced many research issues such as collection development and management, information format, organization, storage, retrieval, search features, display format, user interface, networking, and so on. Advances in network technology, image-processing software, multimedia software coupled with telecommunication technology have improved the various collections of digital libraries in various forms and formats. Some digital libraries have special collections such as multi-media documents, audio, video, images, photographs, maps; while others have combination of general and special collections in digital and print form.

While library automation and Online public Access Catalogues (OPACs) are slowly taking place, digital library research and development activities are at a very preliminary stage in India. Nevertheless, various institutions, especially the universities and other academic institutions in the country have rich resources that can be shared and put to optimum use through digital libraries. In this paper I shall talk about a rather down-to-earth approach to building digital libraries, primarily for the benefit of the academic and research community, but largely for all categories of users in India. Universities and institutions of higher education form the major category of users of digital libraries. In India there are hundreds of universities and other academic institutions catering for the needs of millions of users. INFLIBNET is in an excellent position to spearhead in building such digital libraries, and in this paper I would like to propose some simple guidelines for this.

How Should INFLIBNET Go About Building DLs

Traditional libraries will exist side by side the new and upcoming digital libraries, at least for the foreseeable future. Therefore what we will have are hybrid libraries, a combination of traditional and digital libraries [7,8]. Such hybrid libraries will be extremely useful in developing countries like India, where the objective should be to build digital libraries in niche areas to compliment the traditional library services. External digital library services can be obtained in a number of ways, through subscription, such as through the ACM digital library (http://www.acm.org/dl), or for free such as the NDLTD (http://ww.ndltd.org). Libraries can also provide access to various electronic journals and databases by agreement through the publishers or aggregators/vendors such as the Academic Press

(http://www.idealibrary.com) , EBSCO Online (http://www-sg.ebsco.com/home/default.asp), Ovid online (http://www.ovid.com), and so on, or can provide access to full-text information resources through information services like ProQuest (http://www.umi.com/proquest/). Nevertheless, the country's own resources cannot be accessed and utilized unless a digital library of such resources are built and maintained. There are several areas where such digital libraries can be built, but to my knowledge and understanding the first and foremost objective should be to build a digital library of theses and dissertations. The reasons are obvious. Thousands of theses and postgraduate dissertations produced in Indian Universities and other Institutions for higher learning contain tremendous amount of intellectual contents. These are the products of the intellectual output of the country in many different areas that can not only benefit the academic and research but also can be useful to the government and the commercial world. It is for this reason theses and dissertations have got special priorities in forming the contents of many digital libraries, such as NDLTD(http://www.ndltd.org) and NCSTRL (http://cs-tr.cs.cornell.edu/), etc.

There are different ways INFLIBNET, or any other organization, can go about building a DL of theses and dissertations produced in Indian Universities and other institutions. This paper provides some simple guidelines, that may be easily applied to build DLs on a failsafe basis. The DL of theses and dissertations has to be backed up by a good Document Delivery Service, and fortunately INFLIBNET already has such a service in place (http://www.inflibnet.ac.in/).

Design of a DL of Theses and Dissertations
Theses and dissertations contain the intellectual output of sincere and systematic research carried out by researchers, and monitored as well as reviewed by peers including supervisors and other experts such as examiners, reviewers and editors of journals, when papers are published reporting the research results, and so on. Hence, it is essential that such intellectual output is properly recorded and made available to others in the country and abroad. Several digital libraries have been built throughout the world that provide access to electronic theses and dissertations. Notable examples of digital libraries of theses and dissertations include: NDLTD (http://www.ndltd.org) and NCSTRL (http://cs-tr.cs.cornell.edu/).

DL of Theses and Dissertations in Indian Universities: A Simple Design Approach
One of the most prominent examples of a digital library of theses and dissertations is NDLTD. It is a project initiated and headed by Prof. Edward Fox at Virginia Tech., USA. It is a Federation of Universities and other academic institutions in the US and other parts of the world that have agreed to make their theses and dissertations available in electronic format. While many of the theses and dissertations are available for free, some have restricted access. Any institution can become a member of the Federation, though there are principles to be agreed upon and documents to be signed. While some universities in India can go individually to join such a federation, it may be more appropriate and prudent to build a DL of theses and dissertations in India itself that can be managed easily and can be accessed by national as well as international users, subject to access management principles.

Building a DL containing abstracts of theses and dissertations
Building a digital library of all the theses and dissertations produced at all the Indian Universities and Institutions of Higher learning over the past years (with a cut-off point of, say ten years or so) will no doubt be a massive work. It needs a lot of resources – computational, as well as financial and others. Instead of trying to build one with all the sophisticated features, it is better to take a stage-by-stage approach. In this paper, I would like to propose a simple and failsafe approach; the guidelines are not perfect, and obviously are subject to discussions and modifications, but they would certainly form the foundation.

Stage One
This stage involves some basic and yet very important and critical decisions regarding the nature and structure of the proposed DL, the design and organization issues, access management, technical issues involving hardware, software, and so on. A distributed approach would certainly be useful whereby the participating universities may have their own DL collections that can be accessed by others locally as well as remotely. Adoption of a common simple database design structure would facilitate a seamless access to the digital libraries, held at various universities and institutions, through a common user interface. Design of the DL and selection of software are critical issues, and to begin with I would recommend a simple bibliographic database approach, something similar to the NDLTD but with modifications as necessary, giving search and access facilities to the bibliographic details and abstracts of the theses and dissertations. A simple text retrieval software that is capable of working in the Web environment may be useful. However, a software that can, in future, support full text search capabilities may be an ideal choice. Again, this is a matter of discussions, trials, pilot run, etc.

Developments proposed in stage one above would allow users to search a particular local institution or a collection of regional and/or national institutions, for theses and dissertations and would retrieve a list of selected theses and dissertations with bibliographic details and abstracts. This has to be backed up by an efficient document delivery service. Fortunately INFLIBNET has already have such a service in place that can be used for making the hard copies of theses and dissertations available to the users. The search and retrieval features to be provided will depend upon the software chosen.

Stage two
Once the first phase of the project is put in place and is working, the next obvious stage will be to find ways of making the theses and dissertations available online so that users do not need to go through the expensive and time-consuming document delivery services. The simplest approach would be to make the documents available online in PDF format. PDF files are smaller in size, can contain both text and images, and cannot be altered by users. They may sit on a web server maintained by each participating institution or at a central location that can be retrieved and downloaded by users at the end of their search sessions.

Converting paper documents to a PDF format may be quite a resource-intensive activity, and this cannot possibly be done for all the theses and dissertations produced in the past. A policy framework may be drawn for this purpose to make the selection, if the conversion has to be made. However, all the Universities and institutions may be advised to instruct the researchers to submit a soft copy of their theses that can be stored electronically in PDF format. This may not be a burden for the researchers since most of the theses are now wordprocessed, and all they need is to submit a soft copy of the thesis/dissertation along with the required hard copies. However, in the long run submission of hard copies may be completely unnecessary once everything related to the thesis submission to examination becomes online, as is the case with Virginia Tech, USA (note that the primary objective of their ETD project was to do away with the paper copies of theses and dissertations).

Once the full texts of theses and dissertations are made available online, copyright regulations and access management systems are to be considered to retain and support the intellectual property rights. However, while intellectual forgery may be easy to make (!) in an electronic digital library, at the same time the watchdogs (supervisors, examiners, and the authors themselves) will be better equipped to check for it using the same electronic network.

Stage three
Once the PDF files of theses and dissertations are available on the system, it may be useful to provide full text search capabilities. However, this is a major policy decision, because one has to make a choice between the utility/speed and sophistication of the system. Many digital libraries of today provide full text search capabilities, while a good umber still limit the searches on selected bibliographic fields and abstracts.

Side by side the electronic theses and dissertations, users may need online access to some special collections that may vary from institution to institution and subject to subject. Local digital libraries of special materials may be built and made available to all, or elected users, of the system.

Conclusion
Simple guidelines to building a digital library of theses and dissertations produced in Indian Universities and institutions of higher learning have been proposed in this paper. The guidelines are general and need to be discussed and tested at each stage. Some critical decisions are to be made at some point and these are to be supported by appropriate workshops/ group discussions, systems analysis, pilot testing, etc. Nevertheless, the process of building a digital library of theses and dissertations is rather simple and it can be accomplished without recourse to much resources. The return will however, be immense; research information will be available to the users fast, all round the clock, and this will help them avoid duplication of research efforts and thus wastage of intellectual as well as other resources. Constant access to research information will also stimulate better thinking resulting in new and novel ideas that may eventually benefit the society at large.

References
1. E.A. Fox, The Digital Libraries Initiative: Update and Discussion, *Bulletin of the American Society for Information Science.* October/November (1999) 7-11.

2. Digital Libraries Initiatives. Available at: http://www.dli2.nsf.gov/dlione/ (26 September 2000).

3. eLib: Electronic Libraries Programme. Available at: http://www.ukoln.ac.uk/services/elib/ (visited 25 September).

4. N. Meyyappan, G.G. Chowdhury and Schubert Foo. A review of twenty digital libraries. *Journal of Information Science. 26 (5),* 2000. *331-348*

5. G.G. Chowdhury and Sudatta Chowdhury. An overview of the information retrieval features of twenty digital libraries. *Program,* 34(4), October 2000, 341-373.

6. G.G. Chowdhury and Sudatta Chowdhury. Digital library research: issues and trends. *Journal of Documentation*, 55 (4), 1999, 409—448.

7. C. Rusbridge. Towards the hybrid library. *D-Lib magazine*, 4(7), July/August 1998. Online. Available: http://wwwdlib.org/dlib/july98/rusbridge/07rusbridge.html

8. C. Oppenheim and D. Smithson. What is a hybrid library? *Journal of Information Science,* 25(2), 1999, 97—112.